MATRIX ANALYSIS

矩阵分析〔英〕

ROGER A. HORN
AND
CHARLES R. JOHNSON

Matrix analysis

ROGER A. HORN
The Johns Hopkins University
CHARLES R. JOHNSON
College of William and Mary

The right of the University of Cambridge to print and sell all manner of books was granted by Henry VIII in 1534. The University has printed and published continuously since 1584.

CAMBRIDGE UNIVERSITY PRESS

World Publishing Corp

To the matrix theory community
and
to our families

Dana, Jennifer, and Emily
Susan, Ceres, Corinne, and Howard

for their understanding support

This edition of R.A.Horn & C.R.Johnson—Matrix Analysis
is published by arrangement with Cambridge University Press,
The Edinburgh Building, Shaftesbury Road, Cambridge CB2 2RU.
Licensed for sale in China only. Not for export.
Reprinted by World Publishing Corporation,Beijing,1991
ISBN 7-5062-1029-0

First published 1985
Reprinted with corrections 1987
Reprinted 1988
Reprinted with corrections 1990

Library of Congress Cataloging in Publication Data
Horn, Roger A.
Matrix analysis.
Bibliography: p.
Includes index.
1. Matrices. I. Johnson, Charles R. II. Title
QA188.H66 1985 512.9'434 85-7736
ISBN 0-521-30586-1 hard covers
ISBN 0-521-38632-2 paperback

Preface

Linear algebra and matrix theory have long been fundamental tools in mathematical disciplines as well as fertile fields for research in their own right. In this book, and in the companion volume, *Topics in Matrix Analysis,* we present classical and recent results of matrix analysis that have proved to be important to applied mathematics. The book may be used as an undergraduate or graduate text and as a self-contained reference for a variety of audiences. We assume background equivalent to a one-semester elementary linear algebra course and knowledge of rudimentary analytical concepts. We begin with the notions of eigenvalues and eigenvectors; no prior knowledge of these concepts is assumed.

Facts about matrices, beyond those found in an elementary linear algebra course, are necessary to understand virtually any area of mathematical science, whether it be differential equations; probability and statistics; optimization; or applications in theoretical and applied economics, the engineering disciplines, or operations research, to name only a few. But until recently, much of the necessary material has occurred sporadically (or not at all) in the undergraduate and graduate curricula. As interest in applied mathematics has grown and more courses have been devoted to advanced matrix theory, the need for a text offering a broad selection of topics has become more apparent, as has the need for a modern reference on the subject.

There are a number of well-loved classics in matrix theory, but they are not well suited for general classroom use, nor for systematic individual study. A lack of problems, applications, and motivation; an inadequate index; and a dated approach are among the difficulties confronting readers of some traditional references. More recent books tend to be either elementary texts or treatises devoted to special topics. Our goal was to write a book that would be a useful modern treatment of a broad range of topics.

One view of "matrix analysis" is that it consists of those *topics* in linear algebra that have arisen out of the needs of mathematical analysis, such as multivariable calculus, complex variables, differential equations, optimization, and approximation theory. Another view is that matrix analysis is an *approach* to real and complex linear algebraic problems that does not hesitate to use notions from analysis – such as limits, continuity, and power series – when these seem more efficient or natural than a purely algebraic approach. Both views of matrix analysis are reflected in the choice and treatment of topics in this book. We prefer the term *matrix analysis* to *linear algebra* as an accurate reflection of the broad scope and methodology of the field.

For review and convenience in reference, Chapter 0 contains a summary of necessary facts from elementary linear algebra, as well as other useful, though not necessarily elementary, facts. Chapters 1, 2, and 3 contain mainly core material likely to be included in any second course in linear algebra or matrix theory: a basic treatment of eigenvalues, eigenvectors, and similarity; unitary similarity, Schur triangularization and its implications, and normal matrices; and canonical forms and factorizations including the Jordan form, LU factorization, QR factorization, and companion matrices. Beyond this, each chapter is developed substantially independently and treats in some depth a major topic:

> *Hermitian and complex symmetric matrices* (Chapter 4). We give special emphasis to variational methods for studying eigenvalues of Hermitian matrices and include an introduction to the notion of majorization.
>
> *Norms on vectors and matrices* (Chapter 5) are essential for error analyses of numerical linear algebraic algorithms and for the study of matrix power series and iterative processes. We discuss the algebraic, geometric, and analytic properties of norms in some detail, and make a careful distinction between those norm results for matrices that depend on the submultiplicativity axiom for matrix norms and those that do not.
>
> *Eigenvalue location and perturbation results* (Chapter 6) for general (not necessarily Hermitian) matrices are important for many applications. We give a detailed treatment of the theory of Geršgorin regions, and some of its modern refinements, and of relevant graph theoretic concepts.

Positive definite matrices (Chapter 7) and their applications, including inequalities, are considered at some length. A discussion of the polar and singular value decompositions is included, along with applications to matrix approximation problems.

Component-wise nonnegative and positive matrices (Chapter 8) arise in many applications in which nonnegative quantities necessarily occur (probability, economics, engineering, etc.), and their remarkable theory reflects the applications. Our development of the theory of nonnegative, positive, primitive, and irreducible matrices proceeds in elementary steps based upon the use of norms.

In the companion volume, further topics of similar interest are treated: the field of values and generalizations; inertia, stable matrices, *M*-matrices and related special classes; matrix equations, Kronecker and Hadamard products; and various ways in which functions and matrices may be linked.

This book provides the basis for a variety of one- or two-semester courses through selection of chapters and sections appropriate to a particular audience. We recommend that an instructor make a careful preselection of sections and portions of sections of the book for the needs of a particular course. This would probably include Chapter 1, much of Chapters 2 and 3, and facts about Hermitian matrices and norms from Chapters 4 and 5.

Most chapters contain some relatively specialized or nontraditional material. For example, Chapter 2 includes not only Schur's basic theorem on unitary triangularization of a single matrix, but also a discussion of simultaneous triangularization of families of matrices. In the section on unitary equivalence, our presentation of the usual facts is followed by a discussion of trace conditions for two matrices to be unitarily equivalent. A discussion of complex symmetric matrices in Chapter 4 provides a counterpoint to the development of the classical theory of Hermitian matrices. Basic aspects of a topic appear in the initial sections of each chapter, while more elaborate discussions occur at the ends of sections or in later sections. This strategy has the advantage of presenting topics in a sequence that enhances the book's utility as a reference. It also provides a rich variety of options to the instructor.

Many of the results discussed hold or can be generalized to hold for matrices over other fields or in some broader algebraic setting. However, we deliberately confine our domain to the real and complex fields where familiar methods of classical analysis as well as formal algebraic techniques may be employed.

Though we generally consider matrices to have complex entries, most examples are confined to real matrices, and no deep knowledge of complex analysis is required. Acquaintance with the arithmetic of complex numbers is necessary for an understanding of matrix analysis and is covered to the extent necessary in an appendix. Other brief appendices cover several peripheral, but essential, topics such as Weierstrass's theorem and convexity.

We have included many exercises and problems because we feel these are essential to the development of an understanding of the subject and its implications. The exercises occur throughout as part of the development of each section; they are generally elementary and of immediate use in understanding the concepts. We recommend that the reader work at least a broad selection of these. Problems are listed (in no particular order) at the end of each section; they cover a range of difficulties and types (from theoretical to computational) and they may extend the topic, develop special aspects, or suggest alternate proofs of major ideas. Significant hints are given for the more difficult problems. The results of some problems are referred to in other problems or in the text itself. We cannot overemphasize the importance of the reader's active involvement in carrying out the exercises and solving problems.

While the book itself is not about applications, we have, for motivational purposes, begun each chapter with a section outlining a few applications to introduce the topic of the chapter.

Readers who wish to consult alternate treatments of a topic for additional information are referred to the books listed in the References section following the appendices. These books are cited in the text using a brief mnemonic code; for example, a book by Jones and Smith might be referred to as [JSm]. The codes and complete citations appear alphabetically by author in the References section.

The list of book references is not exhaustive. As a practical concession to the limits of space in a general multitopic book, we have minimized the number of citations in the text. A small selection of references to papers – such as those we have explicitly used – does occur at the end of most sections accompanied by a brief discussion, but we have made no attempt to collect historical references to classical results. Extensive bibliographies are provided in the more specialized books we have referenced. The reader should also be aware of broad and current bibliographical resources covering portions of matrix analysis such as the *KWIC Index for Numerical Linear Algebra* [CaLe] and sections 15 and 65 of the *Mathematical Reviews*.

We appreciate the helpful suggestions of our colleagues and students who have taken the time to convey their reactions to the class notes and preliminary manuscripts that were the precursors of the book. They include Wayne Barrett, Leroy Beasley, Bryan Cain, David Carlson, Dipa Choudhury, Risana Chowdhury, Yoo Pyo Hong, Dmitry Krass, Dale Olesky, Stephen Pierce, Leiba Rodman, and Pauline van den Driessche.

R.A.H.
C.R.J.

Contents

CHAPTER 0

Review and miscellanea

0.0 Introduction

The purpose of this chapter is to catalog briefly, without proof, a number of useful concepts and facts, many of which implicitly or explicitly underlie the material covered in the main portion of the book. Much of this material would be included, in some form, in an elementary course in linear algebra, but we also include a number of useful items that are not commonly found elsewhere or that do not easily fit into the subsequent structure. Thus, this section may serve the reader as a short review prior to beginning the book or as a convenient reference when necessary. We also use this chapter to set basic notation and give some definitions; thus, reference to it will also be useful for these purposes. We do assume that the reader is already familiar with the elementary concepts of linear algebra and with mechanical aspects of matrix manipulations, such as matrix multiplication and addition.

0.1 Vector spaces

Though generally implicitly, and not usually explicitly, involved in the treatment in this book, a vector space is the fundamental setting for matrix theory.

0.1.1 **Scalar field.** Underlying a vector space is the *field,* or set of scalars, from which multiplication occurs. For our purposes, that underlying field will almost always be the real numbers **R** or the complex numbers **C** (see Appendix A) under the usual addition and multiplication, but it could be the rational numbers, the integers modulo a specified prime number, or some other field. When the field is unspecified, we use the symbol $\mathbf{F}$. To qualify as a field, a set of scalars must be closed under two specified binary operations ("addition" and "multiplication"); both operations must be associative and commutative and have an identity element in the set; inverses must exist in the set for all elements under the addition operation and for all elements except the additive identity (0) under the multiplication operation; the multiplication operation must also be distributive over the addition operation.

0.1.2 **Vector spaces.** A *vector space* V over a field $\mathbf{F}$ is a set V of objects (called vectors) which is closed under a binary operation ("addition") which is associative and commutative and has an identity ("0") and additive inverses in the set. The set is also closed under an operation of left multiplication of the vectors by elements of the scalar field $\mathbf{F}$, with the following properties for all $a, b \in \mathbf{F}$ and all $x, y \in V$: $a(x+y) = ax+ay$, $(a+b)x = ax+bx$, $a(bx) = (ab)x$, and $ex = x$ for the multiplicative identity $e \in \mathbf{F}$.

For a given field $\mathbf{F}$, the set $\mathbf{F}^n$ of n-tuples (n a positive integer) with components from $\mathbf{F}$ forms a vector space over $\mathbf{F}$ under the obvious operations (component-wise addition in $\mathbf{F}^n$). The special cases $\mathbf{R}^n$ and $\mathbf{C}^n$ are the basic vector spaces of this book. The set of polynomials with real or with complex coefficients (of no more than a specified degree *or* of arbitrary degree) and the set of real or complex valued continuous functions or arbitrary functions on an interval $[a, b] \subset \mathbf{R}$ are also examples of vector spaces (over $\mathbf{R}$ or $\mathbf{C}$). There is, of course, a fundamental difference between the finite-dimensional space $\mathbf{R}^n$ and the infinite-dimensional vector space of real-valued continuous functions on $[0, 1]$.

0.1.3 **Subspaces and span.** A *subspace* U of a vector space V is a subset of V that is, by itself, a vector space over the same scalar field. For example, $\{[a, b, 0]^T : a, b \in \mathbf{R}\}$ is a subspace of $\mathbf{R}^3$. Usually a subspace of a vector space V is defined by some relation that identifies particular elements of V in such a way that the resulting set is closed under the addition in V – for example, the elements of $\mathbf{R}^3$ with last component 0. It is in this regard that it is useful to think of the resulting set as a subspace rather than as a vector space in its own right. In any event, the intersection of two subspaces is again a subspace.

If S is a subset of a vector space V, the *span* of S is the set Span $S \equiv \{a_1 v_1 + a_2 v_2 + \cdots + a_k v_k : a_1, \ldots, a_k \in \mathbf{F},\ v_1, \ldots, v_k \in S,\ k = 1, 2, \ldots\}$. Notice that Span S is always a subspace even if S is not a subspace. The set S is said to span the vector space V if Span $S = V$.

0.1.4 **Linear dependence and independence.** A set of vectors $\{x_1, x_2, \ldots, x_k\}$ in a vector space is said to be *linearly dependent* if there exist coefficients $a_1, \ldots, a_k$, not all 0, in the underlying scalar field $\mathbf{F}$ such that

$$a_1x_1 + a_2x_2 + \cdots + a_kx_k = 0$$

Equivalently, one of the x_i terms is a linear combination, with coefficients from $\mathbf{F}$, of the others. For example $\{[1,2,3]^T, [1,0,-1]^T, [2,2,2]^T\}$ is a linearly dependent set in $\mathbf{R}^3$. A subset of V that is not linearly dependent over $\mathbf{F}$ is said to be *linearly independent*. For example, $\{[1,2,3]^T, [1,0,-1]^T\}$ is a linearly independent set in $\mathbf{R}^3$. It is important to note that both concepts intrinsically pertain to *sets* of vectors. Any subset of a linearly independent set is linearly independent; $\{0\}$ is a linearly dependent set; and hence any set which includes the 0 vector is linearly dependent. It can happen that a set of vectors is linearly dependent, while any proper subset of it is linearly independent.

0.1.5 **Basis.** A subset S of a vector space V is said to *span* V if every element of V may be represented as a linear combination (with coefficients from the underlying scalar field) of elements of S. For example, $\{[1,0,0]^T, [0,1,0]^T, [0,0,1]^T, [1,0,-1]^T\}$ spans $\mathbf{R}^3$ over $\mathbf{R}$ (or $\mathbf{C}^3$ over $\mathbf{C}$). A linearly independent set which spans a vector space V is called a *basis* for V. Bases are highly nonunique, but are very efficient in that each element of V can be represented in terms of the basis in one and only one way, and this is no longer true if any element whatsoever is appended to or deleted from the basis. An independent set in V is a basis of V if and only if no set which properly contains it is independent. A set that spans V is a basis for V if and only if no proper subset of it still spans V. Every vector space has a basis.

0.1.6 **Extension to a basis.** Any linearly independent set in a vector space V may be extended to a basis of V; that is, given a linearly independent set $\{x_1, x_2, \ldots, x_k\}$ in V, there exist additional vectors $x_{k+1}, \ldots, x_n, \ldots \in V$ such that $\{x_1, \ldots, x_n, \ldots\}$ is a basis of V. The extension of a given independent set to a basis is, of course, not unique [for example, any vector with nonzero third component may be appended to the independent set $\{[1,0,0]^T, [0,1,0]^T\}$ to produce a basis of $\mathbf{R}^3$]. The example of the real vector space $C[0,1]$ of real-valued continuous functions on $[0,1]$ shows that a basis need not, in general, be finite; the infinite set of monomials $\{1, x, x^2, x^3, \ldots\}$ is an independent set in $C[0,1]$.

0.1.7 **Dimension.** If some basis of the vector space V consists of a finite number of elements, then all bases have the same number of elements; this common number is called the *dimension* of the vector space V, and is denoted by dim V. In this event, V is said to be finite-dimensional; otherwise V is said to be infinite-dimensional. In the infinite-dimensional case (e.g., $C[0,1]$), there is a one-to-one correspondence between the elements of any two bases. The real vector space $\mathbf{R}^n$ has dimension n. The vector space $\mathbf{C}^n$ has dimension n over the field $\mathbf{C}$ but has dimension $2n$ over the field $\mathbf{R}$. The basis $\{e_1, e_2, \ldots, e_n\}$ in which e_i has a 1 as its ith component and 0's elsewhere is sometimes called the *standard basis* of $\mathbf{R}^n$ or $\mathbf{C}^n$.

0.1.8 **Isomorphism.** If U and V are vector spaces over the same scalar field $\mathbf{F}$, and if $f: U \to V$ is an invertible function such that $f(ax+by) = af(x)+bf(y)$ for all $x, y \in U$ and all $a, b \in \mathbf{F}$, then f is said to be an *isomorphism* and U and V are said to be isomorphic ("same-structure"). Two finite-dimensional vector spaces over the same field are isomorphic if and only if they have the same dimension; thus, any n-dimensional vector space over the field $\mathbf{F}$ is isomorphic to $\mathbf{F}^n$. Any n-dimensional real vector space is, therefore, isomorphic to $\mathbf{R}^n$, and any n-dimensional complex vector space is isomorphic to $\mathbf{C}^n$. Specifically, if V is an n-dimensional vector space over a field $\mathbf{F}$ with specified basis $\mathcal{B} = \{x_1, \ldots, x_n\}$, then, since any element $x \in V$ may be written uniquely as $x = a_1x_1 + \cdots + a_nx_n$, $a_i \in \mathbf{F}$, $i = 1, \ldots, n$, we may associate x with the n-tuple $[x]_{\mathcal{B}} = [a_1, \ldots, a_n]^T$, relative to the basis $\mathcal{B}$. The mapping $x \to [x]_{\mathcal{B}}$ is an isomorphism between V and $\mathbf{F}^n$ for any basis $\mathcal{B}$.

0.2 Matrices

The fundamental object of study here may be thought of in two important ways: as a rectangular array of scalars and as a linear transformation between two vector spaces, given specified bases for each space.

0.2.1 **Rectangular arrays.** A *matrix* is an m-by-n array of scalars from a field $\mathbf{F}$. If $m = n$, the matrix is said to be square. The set of all m-by-n matrices over $\mathbf{F}$ is denoted by $M_{m,n}(\mathbf{F})$, and $M_{n,n}(\mathbf{F})$ is abbreviated to $M_n(\mathbf{F})$. In the most common case in which $\mathbf{F} = \mathbf{C}$, the complex numbers, $M_n(\mathbf{C})$ is further abbreviated to M_n, and $M_{m,n}(\mathbf{C})$ to $M_{m,n}$. Matrices are usually denoted by capital letters. For example, if

$$A = \begin{bmatrix} 2 & -\frac{3}{2} & 0 \\ -1 & \pi & 4 \end{bmatrix}$$

then $A \in M_{2,3}(\mathbf{R})$. A *submatrix* of a given matrix is a rectangular array lying in specified subsets of the rows and columns of a given matrix.

For example $[\pi \;\; 4]$ is a submatrix (lying in row 2 and columns 2 and 3) of A, above.

0.2.2 **Linear transformations.** Let U be an n-dimensional vector space and V be an m-dimensional vector space over the same scalar field $\mathbf{F}$; let $\mathcal{B}_U$ be a basis of U and $\mathcal{B}_V$ be a basis of V. We may use the isomorphisms $x \to [x]_{\mathcal{B}_U}$ and $y \to [y]_{\mathcal{B}_V}$ to represent vectors in U and V as n-tuples and m-tuples over $\mathbf{F}$, respectively. A linear transformation is a function $T: U \to V$ such that $T(a_1x_1 + a_2x_2) = a_1T(x_1) + a_2T(x_2)$ for arbitrary scalars a_1 and a_2 and vectors x_1 and x_2. A matrix $A \in M_{m,n}(\mathbf{F})$ corresponds to a linear transformation $T: U \to V$ in the following way: The vector $y = T(x)$ if and only if $[y]_{\mathcal{B}_V} = A[x]_{\mathcal{B}_U}$. The matrix A is said to represent the linear transformation T (relative to the bases $\mathcal{B}_U$ and $\mathcal{B}_V$); the representing matrix A depends upon the bases chosen. When we study the matrix A, we realize we are studying a linear transformation relative to a particular choice of bases, but explicit appeal to the bases is usually not necessary.

0.2.3 **Vector spaces associated with a given matrix or linear transformation.** There is no loss of generality in associating an n-dimensional vector space over $\mathbf{F}$ with $\mathbf{F}^n$, and we shall think of $A \in M_{m,n}(\mathbf{F})$ as a linear transformation from $\mathbf{F}^n$ to $\mathbf{F}^m$ (and also as an array). The *domain* of such a linear transformation is $\mathbf{F}^n$; its *range* is $\{y \in \mathbf{F}^m : y = Ax \text{ for some } x \in \mathbf{F}^n\}$. The *null space* of A is $\{x \in \mathbf{F}^n : Ax = 0\}$. The range of A is a subspace of $\mathbf{F}^m$, and the null space of A is a subspace of $\mathbf{F}^n$. We have the relation

$$n = \text{dimension of null space of } A + \text{dimension of the range of } A$$

between these two subspaces.

0.2.4 **Matrix operations.** Matrix addition is defined entry-wise for arrays of the same dimensions and is denoted by + ("$A+B$"). It corresponds to addition of linear transformations (relative to the same basis), and it inherits commutativity and associativity from the scalar field. The zero matrix (all entries zero) is the identity under addition, and $M_{m,n}(\mathbf{F})$ is itself a vector space over $\mathbf{F}$. Matrix multiplication is defined in the usual way, is denoted by juxtaposition, AB, and corresponds to the composition of linear transformations. As such, it is defined only when $A \in M_{m,n}(\mathbf{F})$, $B \in M_{p,q}(\mathbf{F})$, and $p = n$; it is associative. It is not, in general, commutative, for example,

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \neq \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}$$

but it can be commutative when restricted to certain subsets of $M_n(\mathbf{F})$, which are worthy of study. There is an identity under matrix multiplication, the matrix $I \in M_n(\mathbf{F})$ of the form

$$I = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix}$$

This matrix and all scalar multiples of it (called scalar matrices) commute with all other matrices in $M_n(\mathbf{F})$ and are the only matrices which do so. Matrix multiplication is distributive over matrix addition.

We note here that the symbol 0 is used throughout to denote each of the following: the zero scalar, the zero vector (all components equal to the zero scalar), and the zero matrix (all entries equal to the zero scalar). Generally, the context will make clear which it is, so that confusion need not result. We also use the symbol I to denote the identity matrix of any size. If there is potential for confusion, the dimension will be indicated.

0.2.5 **The transpose and Hermitian adjoint.** If $A = [a_{ij}] \in M_{m,n}(\mathbf{F})$, the *transpose* of A, denoted A^T, is that matrix in $M_{n,m}(\mathbf{F})$ whose entries are a_{ji}; that is, rows are exchanged for columns and vice versa. For example,

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}^T = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Of course, $(A^T)^T = A$. The *Hermitian adjoint* A^* of $A \in M_{m,n}(\mathbf{C})$ is defined by $A^* = \bar{A}^T$, where $\bar{A}$ is the component-wise conjugate. For example,

$$\begin{bmatrix} 1+i & 2-i \\ -3 & -2i \end{bmatrix}^* = \begin{bmatrix} 1-i & -3 \\ 2+i & 2i \end{bmatrix}$$

Both the transpose and the Hermitian adjoint [and the inverse to be discussed in (0.5)] obey the *reverse-order law:* $[AB]^* = B^*A^*$ and $(AB)^T = B^TA^T$, assuming the product is defined. For the conjugate of a product, there is no reversing: $\overline{AB} = \bar{A}\bar{B}$. If $x, y \in M_{n,1} = \mathbf{C}^n$, then y^*x is a scalar, and its Hermitian adjoint and complex conjugate are the same; thus, $(y^*x)^* = \overline{y^*x} = x^*y = y^T\bar{x}$.